PROBLEM OF ALTRUISM

Angarika Deb[1], Daniel Smith[2]

1. Department of Cognitive Science, Central European University, Budapest, Hungary.

2. Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, UK.

Synonyms: Evolution of altruistic cooperation; Evolutionary altruism; Psychological altruism.

Definition: We define two 'problems of altruism'. The first is the classic problem of altruism, defined as the issue of how a behavior which decreases an individual's lifetime reproductive success, while helping another individual (or individuals) increase their lifetime reproductive success, can evolve. We also define a 'second-order problem of altruism', where different authors have different conceptions of what does, and does not, constitute 'altruism', including approaches based on kin selection, multi-level selection theory, short-term altruism and psychological altruism.

Introduction

Why would an organism help others at a cost to themselves, without any future reward or pay-off? This, in essence, is the problem of altruism. Despite notions of 'survival of the fittest' and 'nature, red in tooth and claw', seemingly-altruistic acts abound in the natural world: meerkat sentinels famously forego feeding opportunities to guard the troop from predators; many species of mammals, birds and fish engage in 'cooperative breeding', where subordinates sacrifice their own reproduction to help others breed; and humans regularly help others with no apparent benefit to themselves (e.g., giving blood, donating to charity,

engaging in warfare). This altruistic behavior reaches its zenith in eusocial insects, where huge swathes of the colony surrender reproduction altogether to help raise the offspring of a single queen (or handful of queens). Individual somatic cells within a body can also be thought of as altruistic, as only the germ-line (sperm and eggs) is passed on to the next generation: all other somatic cells in the body therefore work altruistically in service of the germ-line.

In all these cases it would not appear to be in an individual's interest to engage in such sacrificial behavior: for instance, a worker bee is not able to pass her genes down to future generations as they are unable to sire offspring of their own. Natural selection is generally believed to maximize individual fitness, yet the existence - and almost ubiquity - of altruism in nature clearly flouts this general rule, and requires special explanation. Darwin appeared acutely aware of this issue, particularly regarding eusocial insects, and in *On the Origin of Species* (1859 pg. 236) wrote:

> "I…will confine myself to one special difficulty, which at first appeared to me insuperable, and actually fatal to my whole theory. I allude to the neuters or sterile females in insect communities: for these neuters often differ widely in instinct and in structure from both the male and fertile females, and yet, from being sterile, they cannot propagate their kind."

Over the next 100 years a solution to this conundrum of how such altruistic behavior can evolve was touched upon by some (including Sewall Wright, Ronald Fisher and JBS Haldane), but was not comprehensively addressed until William Hamilton in the early 1960s published his concept of 'inclusive fitness' (Hamilton, 1964). Rather than taking an organism's individual fitness in terms of personal reproductive success, inclusive fitness partitions fitness into the effect of the focal individual's actions on their own fitness (so-

called *direct fitness*) and the effect of the focal individual's actions on others' fitness (so-called *indirect fitness*), which is weighted by the coefficient of relatedness (*r*) between the individuals. *r* is the probability that individuals share altruistic genes, above baseline levels of the gene(s) in the population. Ignoring rare cases like greenbeards, *r* is primarily due to genealogical relatedness (so *r* approximates 0.5 for full siblings, 0.25 for aunts/uncles, and so on). This focus on genealogical relatedness has earned Hamilton's theory the name 'kin selection'.

Hamilton's formulation presents a simple rule for when altruism can evolve. If *c* is the direct fitness cost to performing an action and *b* is the benefit to others (in terms of increased reproductive success), altruism can evolve when: $br > c$. This is known as Hamilton's rule, and demonstrates how altruistic behaviors - such as sterile worker castes in eusocial insects - can evolve. As a simple example, imagine that individuals only interact with full siblings. Individuals with the altruistic gene increase their partner's reproduction by 3 children at a cost of a decrease in their own reproduction by 1 child. Given these parameters, b=3, r=0.5 and c=1. Therefore, as $br > c$ (3 * 0.5 = 1.5, which is greater than 1), this behavior can evolve, even though it involves a cost to the focal individual's direct fitness. This concept was pithily summarized by Haldane, who quipped that he would lay down his life for two brothers (r=0.5) or eight cousins (r=0.125).

Formal definition of altruism

From this Hamiltonian perspective, altruism is defined as a behavior which causes a decrease in an individual's lifetime direct fitness, but increases the lifetime fitness of others. The lifetime direct fitness consequences to self are therefore negative, while the lifetime fitness

consequences to recipients are positive (-/+; where the first term is the direct fitness effect of the behaviour on the focal individual (*c*), and the second term is the indirect fitness effect of the behaviour on recipient's fitness (*b*)). This is in contrast to mutually-beneficial behavior which has positive lifetime fitness consequences for both parties (+/+). Together, altruistic and mutually beneficial behavior can be grouped together as 'cooperation' (West et al., 2007). Behavior can also be selfish, which means it increases an individual's direct fitness at a fitness cost to others (+/-), or it may be spiteful, in which case behavior decreases both the direct fitness of the focal individual and the fitness of the recipient (-/-; table 1).

Table 1: A classification of social behaviors based on their direct fitness effects on the focal individual (actor) and the fitness consequences for the recipient.

		Fitness effect on recipient (*b*)	
		+ (positive)	- (negative)
Direct fitness effect on actor (*c*)	+ (positive)	**Mutual benefit**	**Selfishness**
	- (negative)	**Altruism**	**Spite**

Note that altruistic behavior does not necessarily require a decrease in personal fitness relative to selfish behavior if interactions are assortative, which occurs when altruists preferentially interact with each other (that is, $r > 0$). Rather, altruism is defined as a decrease in *direct fitness*, rather than a simple decrease in *personal fitness* (which is an individual's total reproductive success, so includes help received from others; note that personal fitness is also known as 'neighbor-modulated fitness'). When $r=0$ altruists will always have lower personal fitness than selfish types, but when $r>0$ altruists may have greater personal fitness than selfish types, given that altruists preferentially interact with and help one another.

Regardless of whether a personal fitness or inclusive fitness approach is adopted, for any value of *r*, altruists will have lower direct fitness than selfish types. As this is all rather abstract, the difference between personal fitness and inclusive fitness is presented in figure 1, while a worked example comparing the two approaches in a simple Prisoner's Dilemma scenario, is presented in figure 2. The essential difference between the two is that inclusive fitness only considers an individual's actions on their own fitness (direct fitness) plus the fitness benefit *given to* others (weighted by relatedness; indirect fitness), while personal fitness considers an individual's actions on their own fitness (direct fitness) plus the fitness benefits *received from* others (indirect fitness). In the personal fitness formulation '*r*' is conceptualized as the probability of correlated interactions between individuals with altruistic genes (rather than strict genetic relatedness), although in most species the main reason for correlated interactions is genetic relatedness (Birch & Okasha, 2015). Despite different conceptualizations regarding whether indirect fitness is classified as help given to others weighted by genetic relatedness (the inclusive fitness approach) or as help received from others and *r* as the probability of correlated interactions (the personal fitness approach), both approaches are (by and large) mathematically equivalent as they both predict that altruism can evolve when $br > c$ (Birch & Okasha, 2015). For additional details of how to measure inclusive fitness, and various pitfalls to avoid, see Grafen (1984).

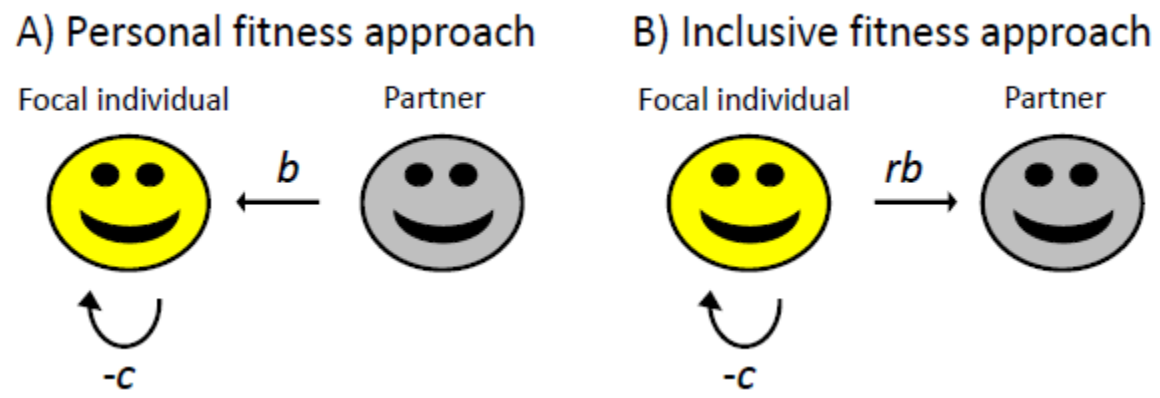


Figure 1: Calculating fitness from: A) the personal fitness approach (also known as 'neighbor-modulated fitness'), in which help (b) is received from others; and B) the inclusive fitness approach, in which help (b) is given to others, weighted by the coefficient of relatedness (r). Note that in both approaches individuals pay an identical cost (c) in direct fitness.

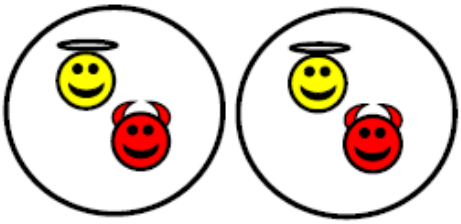
	Random assortment (r=0)	Non-random assortment (r=1)
		
Personal fitness approach	$W_{PF}(A) = w_0 - c$ $= 1 - 1$ $= 0$ $W_{PF}(S) = w_0 + b$ $= 1 + 3$ $= 4$	$W_{PF}(A) = w_0 + b - c$ $= 1 + 3 - 1$ $= 3$ $W_{PF}(S) = w_0$ $= 1$
Inclusive fitness approach	$W_{IF}(A) = w_0 + rb - c$ $= 1 + (0*3) - 1$ $= 0$ $W_{IF}(S) = w_0$ $= 1$	$W_{IF}(A) = w_0 + rb - c$ $= 1 + (1*3) - 1$ $= 3$ $W_{IF}(S) = w_0$ $= 1$

Figure 2: Worked example of how to calculate personal fitness and inclusive fitness given different values of r *(the coefficient of relatedness). The population consists of four individuals, two of which are altruists (smiley faces with a halo) and two of which are selfish types (smiley faces with devil horns). Both strategies begin with a baseline level of fitness* (w_0)*, which has a value of 1. Altruists pay a cost* (c) *of 1 unit of fitness to help others (b), which increases their partner's fitness by 3. The population is split into two groups (*n=*2). If groups form randomly (*r=*0), then groups will be composed of one altruist and one selfish type (left-hand column). While if groups form non-randomly, with altruists always interacting with other altruists (*r=*1), then groups will be composed of either solely altruists or solely*

selfish types (right-hand column). The central row details how to calculate the personal fitness of altruists and selfish types for both r=*0 and* r=*1, while the bottom row details how to calculate inclusive fitness.* $W_{PF}(A)$ *is the personal fitness of altruists,* $W_{PF}(S)$ *is the personal fitness of selfish types,* $W_{IF}(A)$ *is the inclusive fitness of altruists, while* $W_{IF}(S)$ *is the inclusive fitness of selfish types. Although both approaches calculate fitness differently, they both come to the same conclusion; that is, selfish types have higher fitness if groups are formed randomly, while altruists have higher fitness when groups are formed non-randomly (given specific values of* r, b *and* c*).*

Altruism in the real-world

As recognized by Darwin, the behavior of non-reproductive eusocial insects seems to fit this definition of altruism, where individuals help their relatives reproduce, even though they engage in no direct reproduction. Cooperative breeding in vertebrates (such as meerkats, callitrichids, and several bird species) also follows a similar pattern as it often occurs between relatives, so non-reproductive individuals appear to gain indirect fitness benefits by helping their kin. Additionally, as predicted by Hamilton's rule, in insects, birds and mammals, monogamy (and therefore high levels of sibling relatedness) precede the evolution of cooperative breeding (Lukas & Clutton-Brock, 2012).

Despite the empirical and theoretical attention received by indirect benefits as explanations for such seemingly-altruistic behavior, in many cases other potential explanations based on direct fitness benefits may have been overlooked (Clutton-Brock, 2002). For instance, cooperation among kin is not necessarily evidence of altruism if both parties increase their direct fitness, as can occur with reciprocity between relatives. As another example, seemingly-altruistic sentinel behavior among meerkats may in fact be less costly than often assumed, as sentinels are not at an increased risk of predation (Clutton-Brock et al., 1999). Additionally, relatedness to the group does not predict sentinel activity, suggesting that indirect benefits to kin may not explain this behavior. Rather, the best-fed meerkats are more

likely to take up guard duty if a sentinel is not present already, suggesting that this behavior may be largely for selfish reasons to avoid predation.

Furthermore, cooperative breeding in birds - where subordinates help dominants raise the brood - is again often claimed to be due to indirect fitness benefits, as the helpers are often siblings of the chicks. Although these indirect fitness benefits are likely to be important, subordinates may also gain direct fitness benefits by increasing the survival of both themselves and the nest, then subsequently inheriting the breeding rights of the nest after their parents. This appears to explain variation in helping behavior among birds, where the sex which is more likely to inherit the nest provides greater levels of help, despite equivalent levels of relatedness (Downing et al., 2018).

In long-lived animals where all individuals can reproduce, it can be especially difficult to determine whether an organism is acting altruistically, cooperating for mutual benefit, or behaving for purely selfish reasons. Non-reproductive eusocial insects are a rather simplified example as all fitness has to be indirect since reproduction is impossible, but it is much harder to determine whether behavior is altruistic in humans and other animals, as most adults are capable of reproduction and lifetime fitness consequences are hard to measure. Thus, although altruism is easy to define, it may be hard to spot in practice (Grafen, 1984).

Alternative definitions of altruism

The discussion above can be thought of as the 'first-order problem of altruism', which explains the difficulties in altruistic behavior evolving. Next, we discuss the 'second-order problem of altruism' (or 'the problem of the problem of altruism'), in which different authors use different definitions of 'altruism'. This is especially common in the human literature, which is often by necessity cross-disciplinary, meaning that terms common to various

disciplines may have different meanings within each discipline. We discuss three common second-order problems of altruism which are frequently found in the literature. The first concerns different definitions of altruism from a multi-level selection approach, the second concerns altruism measured on different time-scales, while the third concerns conflating evolutionary and proximate definitions of altruism. Although we have presented a definition of altruism derived from inclusive fitness above (which was used as it is the most common definition employed in evolutionary biology; West et al., 2007), we are not advocating that this is the only or correct usage. Rather, we are advocating that whichever decisions authors make regarding the term 'altruism' they are clear on which definition they are using.

Altruism and multi-level selection:

While the definition above based on a kin selection framework is a common definition of altruism in evolutionary biology, it is not the only one that exists in the literature (West et al., 2007). Proponents of a 'multi-level selection' (MLS) perspective on evolutionary change define altruism as a behavior which decreases individual fitness within groups but increases the fitness of the entire group (Sober & Wilson, 1998). The long-term evolutionary trajectory is then determined by the relative strengths of within-group selection (which acts against altruism, as selfish types have greater fitness than altruists within groups) and between-group selection (which favors the spread of altruism, as groups with more altruists have greater fitness than groups with fewer altruists). Note also that although kin selection and MLS approaches appear rather different, they are equivalent approaches mathematically (Birch & Okasha, 2015); they just partition fitness in different ways (kin selection into direct and indirect fitness effects; MLS into within- and between-group components). One view is therefore not necessarily more correct than the other: the difference between these

frameworks is in perspective, not process. While kin selection approaches define altruism in absolute terms over the entire population, from an MLS perspective altruism is defined relative to the group. As will be discussed below, this can lead to different expectations of when altruism can evolve, depending on whether one adopts an inclusive/personal fitness or MLS perspective.

Within the MLS framework there are two flavors of altruism: weak and strong (Wilson, 1990). Weak altruism occurs when altruists have lower fitness than selfish types within groups, but have greater fitness than if they were not altruistic. An example is riding a tandem bike: if one person pedals and the other does not, then the lazy rider exerts less energy than the individual who pedals (i.e., selfish types have greater fitness within groups), but the altruist has greater fitness than if they did not pedal at all, as at least if one person pedals they will reach their destination (Kerr & Godfrey-Smith, 2002). This scenario is presented abstractly in figure 3a. As the altruist increases their own fitness by cooperating, this behavior is not altruistic from a kin selection perspective, but rather may be mutually-beneficial to both parties, or simply self-interested, as the benefit derived by others is an incidental by-product. As such, weak altruism does not require assortment, so this behavior can evolve in randomly-formed groups (i.e., if r=0).

In contrast, strong altruism occurs when altruists possess both lower fitness than selfish types within groups and lower fitness than if they were not altruistic. The example presented in figure 2 of a classic Prisoner's Dilemma is an example of strong altruism (see also figure 3b). If group formation is random, then altruists have lower fitness compared to if they did not cooperate. Thus, if a behavior is strongly altruistic from an MLS perspective, then it is also altruistic from a kin selection perspective, and thus requires assortment to evolve (i.e., r>0).

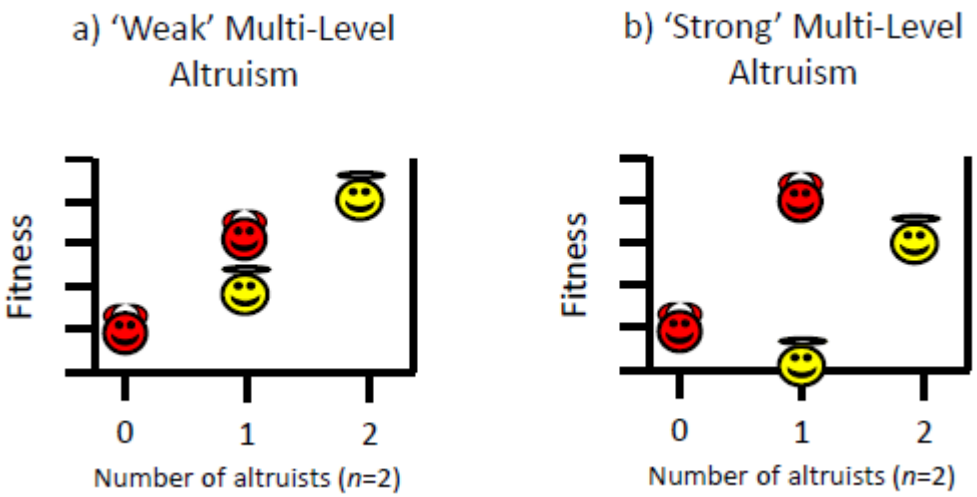


Figure 3: Strong and weak altruism, as defined from a multi-level selection perspective. The population consists of altruists (smiley faces with a halo) and selfish types (smiley faces with devil horns). A) 'Weak' multi-level altruism: Altruists have lower fitness than selfish types within groups (i.e., in a mixed group, selfish individuals have greater fitness than altruists). However, given the choice between being an altruist or being selfish, individuals have greater fitness if they decide to be an altruist (as the fitness of an altruist in a group of one altruist and one selfish type is greater than the fitness of a selfish type in a group of no altruists). B) 'Strong' multi-level altruism: As with weak multi-level altruism, altruists have lower fitness than defectors within groups. However, given the choice between being an altruist or being selfish, individuals have greater fitness if they decide to be selfish (as the fitness of an altruist in a group of one altruist and one selfish type is lower than the fitness of a selfish type in a group of no altruists). Weak altruism does not require assortative group formation in order to occur, while strong altruism does.

One must therefore be careful when converting between altruism as defined from an MLS perspective (which often uses the multi-level Price equation, which partitions selection into between- and within-group components; Sober & Wilson, 1998) and altruism as defined from a kin selection perspective (which uses Hamilton's rule of $br > c$). Under conditions of weak altruism, behavior may be altruistic from an MLS perspective, but not altruistic from a kin selection approach, as acting 'altruistically' increases an individual's direct fitness. Of course, which definition people use is subjective - and a plurality of approaches can help to understand problems in greater detail (Kerr & Godfrey-Smith, 2002). However, it can cause confusion if different definitions of 'altruism' are used synonymously (West et al., 2007).

Short-term altruism:

Some authors define altruism in terms of short-term costs, rather than in terms of life-time direct fitness (e.g., Fehr & Fischbacher, 2003). For instance, in reciprocal altruism, individuals pay a short-term cost to cooperating, but it is not necessarily altruistic in the Hamiltonian sense defined above, as they increase their direct fitness in repeated cooperative interactions. That is, the behavior is mutually-beneficial (+/+), rather than altruistic (-/+; although note that reciprocity is weakly altruistic from an MLS perspective, as groups of reciprocators outcompete selfish groups, even though in mixed groups selfish types have greater fitness than reciprocators; Sober & Wilson, 1998). Several models of 'altruistic punishment' or 'strong reciprocity' also appear to share this feature, where altruism is defined in terms of short-term pay-offs, rather than in terms of lifetime direct fitness (e.g., Gintis, 2000). These features, combined with a focus on between-group competition to spread such 'altruistic' behavior, can make it difficult to determine whether the behavior in these models is altruistic from a kin selection perspective or not (Keller & Lehmann, 2006).

Evolutionary and psychological altruism:

Up until now, we have been dealing with why altruism might have evolved, and how it may be adaptive in a social environment. However, as with everything else in biology, the study of a behavioral trait concerns two distinct questions – *why* a behavior evolved, and *how* the behavior works, i.e., the ultimate and proximate explanations. It is crucial to maintain a separation between these explanations when attempting to understand and explain any trait. For instance, knowing that individuals feel reward when they punish defectors is a proximate answer for why humans punish, but it does not answer why the behavior evolved. An answer to this question would be an ultimate explanation in terms of individuals who detect and

punish free-riders have greater fitness than those who do not detect and punish free-riders. In the case of altruism, evolutionary explanations deal with fitness considerations of individuals and groups, while psychological explanations are largely based on the *motives* of the actor.

We now look at altruism from a psychological perspective. Hamilton's theory postulated that altruistic behavior can evolve due to the inclusive fitness benefits it bears for the actor via indirect reproduction in kin. For these benefits to play a mechanistic role in behavior however, they need to be experienced by the actor in the form of feelings, motivational factors, or moral factors (Monroe, 1994). Psychological altruism deals with the understanding of such motivations that can drive one to benefit others at a cost to self (although unless directed towards kin, such psychologically altruistic behavior is generally framed as being short-term altruistic, rather than evolutionarily altruistic). Empathy, for example, is an emotional mechanism that can propagate other-regarding behavior, by creating a feeling of connectedness (De Waal, 2008). By the psychological mechanisms of emotion contagion and perspective-taking, individuals can 'put themselves in other's shoes' and act in the interests of others and against narrow short-term self-interest. Morality forms a cultural norm, as well as part of one's conscience, that can promote such seemingly-altruistic behavior by dictating that one must help others, sometimes even at a cost to themselves. To the extent that these norms become internalized, other-regarding preferences - and therefore psychological altruism - can emerge.

Though ultimate evolutionary explanations do not consider the internal motives of an actor to determine altruism - after all, eusocial insects are evolutionarily altruistic yet we would be unlikely to ascribe to them altruistic motives - the literature largely persists in using terms that are reflective of these motives. Actions are called "selfish" or "altruistic" based on whether the actor only seeks benefits for herself, or intends to benefit another. This oftentimes ensues confusion between evolutionary and psychological altruism, and leads to

conflation of terms. Since these terms also have high vernacular use – which are mostly to do with one's motives – they can cause further confusion in the evolutionary understanding of altruism (Sober & Wilson 1998). Some authors use evidence of psychological altruism to claim that some human behavior is evolutionarily altruistic. For instance, Richerson and Boyd (2005, pages 216-221) review evidence for empathy and seemingly-altruistic cooperation in one-shot economic games with strangers (to be discussed below) and conclude that humans possess other-regarding preferences. They then use this evidence for psychological altruism to argue that evolutionary altruism towards non-kin may also exist. However, as a general rule there is no clear relationship between evolutionary and psychological definitions of altruism, and they should not be confused with one another, or used interchangeably. Behaviors that are psychologically altruistic, can be evolutionarily either selfish or altruistic, and vice versa.

An example to demonstrate this is a man saving a drowning (unrelated) child. This behavior may decrease the man's lifetime direct fitness (and increase the fitness of the child) and therefore be evolutionarily altruistic. Alternatively, although this behavior involves a short-term cost, this behavior may increase his direct fitness if he is more likely to be chosen as a future reciprocal trading partner or as a mate, in which case this behavior is not evolutionarily altruistic. Similarly, the man's intentions may be psychologically altruistic if he genuinely cares about the child's welfare. However, his motives may not be altruistic if he perceives the scenario as a chance to enhance his reputation, rather than caring about the child. Thus, to classify an act as psychologically altruistic, the actor must *intend* to bring benefits to the recipients, even at a cost to themselves. But definitions of altruism based on intentions are separate from definitions of altruism based on whether behavior is detrimental to evolutionary fitness.

Does psychological altruism exist?

While determining an actor's true intentions and motivations is a difficult task, many psychologists and economists often take a rather pessimistic view towards psychological altruism, in which altruistic acts are treated as a subtle variant of self-interest. For instance, altruistic acts have been explained from an egoistic perspective, in which actors help others solely in order to reap future benefits, for instance by forming reciprocal relationships. Alternatively, these acts have been explained from an egocentric perspective, in which actors behave altruistically if watching the pleasure of her beneficiaries exceeds the satisfaction of consuming the commodity itself (Khalil, 2004). According to both these views, no act is truly psychologically altruistic since the actor gets either reputational advantages, or feels pleasure and internal gratification from their seemingly-altruistic acts (altruistic hedonism). This pessimism has been well captured by Ghiselin (1974) in the following lines, ''What passes for cooperation turns out to be a mixture of opportunism and exploitation...Scratch an altruist, and watch a hypocrite bleed''. Based on these motivations, economists sometimes partition altruism into pure and impure altruism; in pure altruism, the only motive of the actor is the outcome of the altruistic act itself (e.g., aid being given to children in need), whereas in impure altruism, the impact on the recipient's welfare is not the primary motivator, as instead the actor gets a 'warm-glow' or feeling of goodness, which comprises a private benefit (Andreoni 1990).

However, despite these difficulties of determining whether an act is psychologically altruistic or not, some experiments have indicated that certain aspects of human behavior are indeed motivated by a concern for others' welfare. This optimistic view was endorsed by Adam Smith (1759) who wrote that,

> "How selfish soever man may be supposed, there are some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it."

The empathy-altruism hypothesis (Batson et al, 1987) claims that empathy evokes motivation to reduce the other's need, promoting altruism, and that self-benefits from such acts are simply unintended consequences of the ultimate goal (helping the other). In a series of experiments conducted by Piliavin, Dovidio, Gaertner and Clark (1981), there were various degrees of emotional responses in reaction to seeing someone in need. Piliavin and colleagues quantitatively described a low magnitude emotional response to other's needs in non-emergency situations as empathy, such as feeling sympathetic, compassionate or tender, and a high magnitude emotional response to other's needs in emergency situations as personal distress, such as feeling alarmed, upset or disturbed. These emotional and psychological responses to other's needs are hypothesized to provide the motivation for other-regarding behavior in humans (Batson et al, 1987). These and other experiments suggest that humans, at least in part, are motivated by a concern for the welfare of others, indicating that psychological altruism may indeed exist.

Case study of human altruism: Behavior in economic games

Numerous experiments have demonstrated that humans behave seemingly-altruistically in experimental economic games, where the optimal strategy is to free-ride on the cooperation of others: in all societies tested, humans frequently cooperate with strangers, contribute to public pools, punish selfish individuals at a personal cost in one-shot interactions, and even punish those who defect on others (so-called third-party punishment; Henrich et al., 2005; Bernhard et al., 2006). This behavior appears both evolutionarily altruistic - as individuals

help others at a cost to self, despite a lack of repeated interactions or reputational effects - and psychologically altruistic - as individuals appear motivated to assist others and enhance their welfare.

However, leaving aside the question of whether behavior in these games is psychologically altruistic, the assumption that such behavior is detrimental to fitness, and therefore evolutionarily altruistic, has been questioned on several fronts. First, although this behavior is seemingly-altruistic in the context of the game (and is certainly short-term altruistic), the strategies adopted by players in these games may be mutually-beneficial when applied outside the lab in real-world settings. The costs of mistaking repeated interactions for a one-shot interaction may be high, in which case seemingly irrational cooperation in one-shot encounters may evolve (Delton et al., 2011). Second, the design of these experimental games is likely to be unfamiliar to most participants, which mean that individuals tend to use behavioral strategies from outside the lab (which are likely to be more cooperative than the optimal behavior in these games) when first playing these games. Supporting this interpretation are findings that individuals with greater experience in these games are less cooperative than 'naive' participants, presumably because they have learned that it pays not to cooperate in these situations (Rand et al., 2014). Third, it is very difficult to know exactly what behavior in these games means in terms of real-world cooperation, as the external validity of behavior in these games is often low (Gurven & Winking, 2008); although substantial levels of cooperation towards strangers are observed in these experiments, in the real-world unsolicited cooperation towards strangers is much rarer (Winking & Mizer, 2013). Of course, perhaps behavior in these games (and some behavior in the real-world) are examples of evolutionary altruism, but it is very difficult to know definitively, and these alternative interpretations cast doubt on the conclusion that some cooperative behavior in humans (at least towards non-kin) is altruistic.

Conclusion

The concept of altruism has a fraught intellectual history, with numerous varying and conflicting definitions in use. With the term ‘altruism’ being used across disciplines, like evolutionary biology, economics, psychology and philosophy, it is impossible to classify altruism by a single agreed-upon definition. This plurality of approaches means that authors - especially those engaging in cross-disciplinary work - need to clearly define what they mean by ‘altruism’, in order to forestall misunderstandings and unnecessary confusion. This is especially pertinent for the distinction between evolutionary and psychological altruism. In this chapter we have attempted to describe various definitions of altruism found in the literature, including Hamiltonian altruism, multi-level selection altruism (both weak and strong), short-term altruism, and psychological altruism, as well as the links - or lack thereof, in the case of evolutionary and psychological altruism - between these definitions. Given this multitude of definitions, it is no wonder that, despite being a topic of interest for centuries, we still do not have an established understanding and consensus on ‘what is altruism’ and whether much human behavior is in fact altruistic (both from evolutionary and psychological perspectives). Clarifying one’s own meaning of altruism can go a long way to solve this second-order problem of altruism, and allow coherent interaction between disciplines in order to understand whether, and under what circumstances, altruism may evolve.

Cross-references

Altruism among Nonkin; Reciprocal Altruism and Cooperation for Mutual Benefit; Reputation and Altruism; Psychology of Reciprocal Altruism; Altruism Norms; Evolution of Reciprocal Altruism; Kin Selection Hypothesis; C<Rb; Cooperation varies with Genetic Relatedness; Hamilton's Rule; Hamilton's Rule and Theoretical Implications; Hamilton's Rule and Kin Investment; Altruistic Punishment and Strong Reciprocity; High-Cost Altruistic Helping; Group Selection; Multilevel Selection Theory.

References

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods : A Theory of Warm-Glow Giving. *The Economic Journal*, *100*(401), 464–477.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Adults' emotional reactions to the distress of others. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 163–185). Cambridge: Cambridge University Press.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*(7105), 912–915. https://doi.org/10.1038/nature04981

Birch, J., & Okasha, S. (2015). Kin selection and its critics. *BioScience*, *65*(1), 22–32. https://doi.org/10.1093/biosci/biu196

Clutton-Brock, T. (2002). Breeding together: kin selection and mutualism in cooperative vertebrates. *Science (New York, N.Y.)*, *296*(5565), 69–72. https://doi.org/10.1126/science.296.5565.69

Clutton-Brock, T. H., Riain, M. J. O., Brotherton, P. N. M., Gaynor, D., Kansky, R., Griffin, A. S., & Manser, M. B. (1999). Selfish Sentinels in Cooperative Mammals. *Science*, *284*, 1640–1644. https://doi.org/10.1126/science.284.5420.1640

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

de Waal, F. B. M. (2008). Putting the Altruism Back into Altruism: The Evolution of Empathy. *Annual Review of Psychology*, *59*(1), 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters.

Proceedings of the National Academy of Sciences, *108*(32), 13335–13340. https://doi.org/10.1073/pnas.1102131108

Downing, P. A., Griffin, A. S., & Cornwallis, C. K. (2018). Sex differences in helping effort reveal the effect of future reproduction on cooperative behaviour in birds. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1885), 20181164. https://doi.org/10.1098/rspb.2018.1164

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791. https://doi.org/10.1038/nature02043

Ghiselin, M. T. (1974). *The economy of nature and the evolution of sex*. Berkeley: University of California Press.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*(2), 169–179. https://doi.org/10.1006/jtbi.2000.2111

Grafen, A. (1984). Natural Selection, Kin Selection and Group Selection. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology* (pp. 62–84). Oxford: Blackwell Scientific Publications.

Gurven, M., & Winking, J. (2008). Collective action in action: Prosocial behavior in and out of the laboratory. *American Anthropologist*, *110*(2), 179–190. https://doi.org/10.1111/j.1548-1433.2008.00024.x

Hamilton, W. D. (1964). The Genetical Evolution of Social Behavior. *Journal of Theoretical Biology*, *7*, 17–52. https://doi.org/10.1016/0022-5193(64)90039-6

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *The Behavioral and Brain Sciences*, *28*(6), 795–815. https://doi.org/10.1017/S0140525X05000142

Kerr, B., & Godfrey-Smith, P. (2002). Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy*, *17*(4), 477–517. https://doi.org/10.1023/A:1020504900646

Khalil, E. L. (2004). What is altruism? *Journal of Economic Psychology*, *25*, 97–123. https://doi.org/10.1016/j.tree.2003.10.004

Lehmann, L., & Keller, L. (2006). The evolution of cooperation and altruism - A general framework and a classification of models. *Journal of Evolutionary Biology*, *19*(5), 1365–1376. https://doi.org/10.1111/j.1420-9101.2006.01119.x

Lukas, D., & Clutton-Brock, T. (2012). Cooperative breeding and monogamy in mammalian societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1736), 2151–2156. https://doi.org/10.1098/rspb.2011.2468

Monroe, K. R. (1994). A Fat Lady in a Corset: Altruism and Social Theory. *American Journal of Political Science*, *38*(4), 861–893.

Piliavin, J. A., Dovidio, J. F., Gaertner, S. L., & Clark, R. D. I. (1981). *Emergency Intervention*. New York: Academic Press.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677. https://doi.org/10.1038/ncomms4677

Richerson, P., & Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Smith, A. (1759). *The Theory of Moral Sentiments*. London: A. Miller.

Sober, E., & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*(2), 415–432. https://doi.org/10.1111/j.1420-9101.2006.01258.x

Wilson, D. S. (1990). Weak Altruism, Strong Group Selection. *Oikos*, *59*(1), 135–140.

Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, *34*(4), 288–293. https://doi.org/10.1016/j.evolhumbehav.2013.04.002